The Effect of Intonation on Children's Perception of Contrast in Visual Search

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation
with research distinction in Linguistics in the undergraduate colleges
of The Ohio State University

by
Sarah Alaine Bibyk

The Ohio State University
June 2010

Project Advisor: Professor Shari Speer, Department of Linguistics

Abstract

The acquisition of prosody is an area that has been the subject of much debate, particularly with regard to the acquisition of pitch accents. In English a pitch accent is a prominent local excursion in the tune of an utterance described as being low, high, or some combination of the two and which is used to convey pragmatic information about the utterance. Much of previous research investigating children's understanding of specific pitch accents has concluded that children have a poor grasp of them—even well into grade-school ages (Cruttenden, 1985; Cutler & Swinney, 1987; Gualmini, Maciukaite, & Crain, 2002; Hornby, 1971; Wells, Peppé, & Goulandris, 2004). However, more recent research with Japanese-speaking 6-year-olds has shown that children are able to comprehend contrastive focus and even use it to make predictions about upcoming speech (Ito, Jincho, Minai, Yamane, & Mazuka, ms; Ito, Jincho, Yamane, Minai, & Mazuka, 2009a & 2009b). This research was adapted in order to measure English-speaking 6- to 7-year-old children's responses to the L+H* pitch accent while the children were engaged in a visual search task. This research found that children between the ages of 6 and 7 do comprehend the implications of L+H* and in fact show facilitated fixations to a target when L+H* is used felicitously and erroneous fixations when L+H* is used infelicitously. The timing of the children's fixations however were delayed as compared to the adult control group, which suggests that children's processing of these pitch accents is still under development between the ages of 6 and 7.

English prosody

The question as to how and when children acquire the prosodic system of their language is by no means trivial. Like grammar, prosody has a hierarchical structure with rules for combining the units into well-formed tunes (Beckman, 1996). At a first glance, a tune is the

fundamental frequency pattern produced by the human vocal chords during speech. In reality, however, various other factors (such as the overall stress pattern and syllable loudness) affect how we control pitch (Pierrehumbert 1980). In the theory of prosody outlined by Pierrehumbert (1980) tunes are composed of a series of high (H) and low (L) tones as defined by the pitch range of each individual speaker. Every utterance, or intonational phrase, has a tune. Each intonational phrase in turn is composed of at least one intermediate phrase and each intermediate phrase must contain at least one pitch accent. In addition, intonational phrases must end with a boundary tone which can be H or L. Similarly intermediate phrases must have a H or L phrasal tone. It follows then that even single word utterances are intonational phrases, and therefore must also be an intermediate phrase with a pitch accent, a phrasal tone, and a boundary tone.

The alignment of the prosodic structure to the words of an utterance is defined by the stress pattern of that utterance. In English, stress is essentially the relative prominence of individual syllables (Beckman & Pierrehumbert, 1986). Typically syllables that are perceived as stressed have acoustically measurable differences from those perceived as unstressed such as longer syllable duration and hyper-articulation of the syllable's segmental content. An example of this difference is the noun ob'-ject as compared to the verb ob-ject': the diacritic ' indicates the stressed syllable. Notice that the first syllable of the verb is shorter in comparison to the first syllable of the noun. The first syllable of the verb has also undergone what is known as vowel reduction (a transition from the low back vowel to the medial vowel schwa). There do exist in English a few minimal pairs where the only difference between the pair is the stress pattern (compare in'-sight versus in-cite'), which is to say the stressed syllables show differences in intensity and duration but no articulatory differences in the vowels (Cutler, 1986). These,

however, are exceedingly rare (Cutler, 1986). Predominantly, stressed syllables are longer and more fully articulated and it is these syllables to which pitch accents are aligned.

In the prosodic transcription system ToBI (Tones and Break Indices: Beckman & Ayers, 1997) English has five distinct pitch accents: two simple (H* and L*) and three complex (L*+H, L+H*, H+!H*). In each of these accents the letters indicate whether the tone is high or low and the * indicates which part of the accent is aligned to the word's stressed syllable. For example with H* (called *high star*) there is only one tone so this is the tone aligned with the stressed syllable. In comparison, L+H* (called *low plus high star*) has two tone targets: an initial low target followed by a high target which is aligned to the stressed syllable. The diacritic ! in the pitch accent H+!H* indicates a process in speech known as "downstepping" where by the compression of the pitch range a subsequent H tone is slightly lower as compared to the initial H tone (Beckman & Ayers, 1997). This research will only address H* and L+H* with the primary focus on children's understanding of the latter.

The meaning of H and L+H**

Pierrehumbert and Hirschberg (1990) attempted to address how prosody contributes to discourse interpretation by detailing the meanings assigned to the various pitch accents and how those meanings combine with the meanings of different phrasal and boundary tones. The purpose of H* in their theory is to highlight items which are "new" with respect to the discourse. When H* is combined with a L phrasal tone (indicated by the diacritic -) and a L boundary tone (indicated by the diacritic %), it forms what Pierrehumbert and Hirschberg (1990) called the "neutral declarative" intonational contour. The function of this contour is simply to convey information as in

1. The train leaves at seven

H* H* H* L-L%

L+H*, in comparison, evokes a scale which has different interpretations depending upon the context. In particular Pierrehumbert and Hirschberg (1990) claimed L+H* is used to convey that "the accented item—and not some alternative item—should be mutually believed" in a discourse. They also noted that L+H* is commonly used to make a correction or a contrast between two items. In the following example

2. I didn't want vanilla cake. I wanted CHOColate cake.

L+H* L-L%

the L+H* accent on the stressed syllable of "chocolate" establishes a contrast between the two types of cake, vanilla and chocolate. If the L+H* were shifted however to the word "cake," the contrast would be with some other type of dessert as in

3. I didn't want chocolate ice cream. I wanted chocolate CAKE.

L+H* L-L%

The mere presence of L+H* in an utterance, however, does not necessarily evoke a contrastive interpretation. Ito and Speer (2008) found that when L+H* is used to accent a discourse marker, it does not prime a contrastive interpretation between items mentioned in that same utterance. In their experiment they instructed participants to decorate small Christmas trees with ornaments of different shapes and colors and the order of the ornaments was controlled in order to elicit specific contrasts. They found that in sequences such as the example below

4. Hang the blue ball. Now, hang the GREEN ball.

L+H*

if the color adjective of the second ornament was accented with L+H*, participants looked sooner to the target object as compared to when the adjective was accented with H*. They interpreted this as an indication that the participants were aware of the contrastive meaning of L+H* and used it to make a prediction about the upcoming referent. However if the L+H* instead appeared on the discourse marker used to transition between the two utterances

5. Hang the blue ball. NOW, hang the green ball.

L+H*

the participants showed no anticipatory looks to the target.

There is some dispute as to what the "meaning" of specific pitch accents like L+H* and H* are. The definitions established by Pierrehumbert and Hirschberg (1990) and assumed by the ToBI labeling system (Beckman & Ayers, 1997) are not universally accepted among researchers. Ladd and Schepman (2003) for example claimed that L+H* is merely a variant of H*. They noted pitch accents transcribed as H* often also follow a preceding low target which is not aligned to any syllable.

However, how the accents are labeled is largely irrelevant for how these accents are interpreted by adults in speech. Recent research by Watson, Tanenhaus, and Gunlogson (2008) found that adults anticipate L+H* to be evoking a contrast, whereas they assume H* will refer to either contrasted or new items, suggesting that it is to some extent "unmarked" or neutral. They used the visual world paradigm developed by Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995) and used by various researchers to investigate the effect of pitch accent on speech processing (Dahan, Tanenhaus, & Chambers, 2002; Weber, Braun, & Crocker, 2006; Ito & Speer 2008). They monitored their participants' eye-movements as the participants were instructed to move objects around on a computer screen. The target objects varied in terms of whether or not

they were new or given relative to the experimental context of the discourse, and also whether or not they were part of a contrast set. The target objects also varied in whether they were produced with a H* or a L+H* accent. Watson et al. found that adults looked more at a member of the contrast set when they heard L+H* as compared to H*, and furthermore that adults looked more to a discourse new item when they heard H* as compared to L+H*. Importantly, their results show that H* and L+H* are treated as different categories by adults, which is one of the fundamental assumptions made by the experimental design of this research.

While pitch accents do have correlates in the physical world such as durational and fundamental frequency contour differences, the interpretation of these correlates is highly dependent upon the intonational context and even on each speaker's individualized pitch range. A woman's productions of L+H*, for example, are physically quite different from a man's productions. Children acquiring English are thus tasked with associating physical properties like duration and fundamental frequency with abstract meaning as defined by the intonational structure of their language in order to produce and comprehend English pitch accents correctly.

Previous child research

There is much disagreement in the literature on children's acquisition of pitch accents, in particular with regard to when children acquire adult-level competence in comprehension of pitch accents. Most researchers claim that acquisition is late, in some cases not even until the age of 10 (Cruttenden, 1985; Cutler & Swinney, 1987; Gualmini, Maciukaite, & Crain, 2002; Hornby, 1971; Wells, Peppé, & Goulandris, 2004). Perplexingly children performed well in tasks testing their production of pitch accents (Hornby & Hass, 1970; Macwinney & Bates, 1978; Wells et al. 2004) leading researchers to conclude that production must precede comprehension with regard to pitch accents.

However, current researchers (Ito et al., ms, 2009a & 2009b) have speculated that children's seemingly poor comprehension of pitch accent may in fact be largely due to the nature of the tasks designed to test them. Most experiments relied on some type of offline measurement; comprehension was not assessed directly but instead by means of some secondary action performed by the participant. For example, Cruttenden (1985) presented children with verbal utterances such as "John's got FOUR oranges" and asked children to match this sentence to one of three pictures: 1. a boy with four oranges and a girl with two oranges, 2. a boy with four oranges and a girl with four bananas, or 3 a boy with three oranges and a girl with four oranges. In this case the correct answer is the first picture. Cruttenden found that even children as old as 10 did not select the correct picture more than chance. Similarly, Wells et al. (2004) played the children prerecorded sentences such as "I wanted CHOCOLATE and honey" and then asked the children them to indicate which item the speaker did not receive based on the sentence. In this example the correct answer is "chocolate." Wells et al. found that only the 13-year-old age group was able to choose the correct answer more times than chance. Not only do these experiments involve an indirect measurement for ascertaining children's comprehension of accent, but the tasks also require that children reconstruct the context which licenses the proper use of the pitch accents since the sentences were presented in isolation. For example, in order for children answer correctly in Cruttenden's (1985) task, children would need to posit an utterance such as "Jill's got two oranges" after which "But John's got FOUR oranges" could legally follow. To properly measure children's language comprehension abilities the task must provide a rich enough context (i.e. provide utterances explicating the context) in order to license the use of L+H*. It must also use some type of online measurement.

Cutler and Swinney (1987) attempted to measure children's processing of pitch accents by using reaction time as a measure. In one particular experiment they instructed children to monitor for target words in prerecorded sentences and to press a response button when they heard one of the target words. In certain conditions the target words were "accented" (produced with greater intensity and duration). Cutler and Swinney's prediction was that accented words would produce faster reaction times as compared to when the target words were unaccented. Cutler and Swinney were able to find a significant difference in reaction times for accented and unaccented words in 6-year-olds (accented words produced fasted reaction times), but 5-year-olds showed no difference in reaction time between accented and unaccented words, leading Cutler and Swinney to conclude that the children were not yet adult-like in their ability to perceive accent. However this experiment did not test whether or not children understand the implications of pitch accents in context since Cutler and Swinney were using the accentation to "highlight" particular words in isolated sentences. Thus their results only show that 6-year-old children react faster to words that are more acoustically salient, not that the 6-year-olds were aware of the contrastive meaning of L+H*. To truly test children's abilities to comprehend pitch accents, the children must be given a task which explicitly uses the pitch accents in context.

Eyetracking Studies

A technique that has met with much success in measuring human processing of prosody is eyetracking. In most eyetracking experiments participants are seated in front of either a large screen which is used to present a visual display (Dahan et al., 2002; Gennari, Meroni, & Crain, 2005; Weber et al. 2006) or in front of an actual display of real objects (Ito & Speer, 2008, Snedeker & Trueswell, 2003). A camera is used to record the eye-movements of the participants as they receive verbal instructions to complete some kind of task with the displayed objects.

Eye-movements have been shown to be an effective means of accessing real-time processing of spoken language, because listeners' attention, and thus their gaze, is drawn to displayed objects when they are named in the spoken instructions (Trueswell & Tanenhaus, 2005). Some experimenters have used an eyetracking technique to show that adults use pitch accent in the interpretation of syntactically ambiguous sentences that are disambiguated by the placement of the accent (Gennari et al., 2005). Other researchers have shown that adults use pitch accents make predictions about the content of upcoming speech (Dahan et al., 2002; Ito & Speer, 2008; Snedeker & Trueswell, 2003; Weber et al., 2006). Dahan et al. (2002), for example, presented participants with displays that contained four objects, two of which overlapped in terms of their onset and thus formed a contrast pair (candle and candy). In the critical trials participants heard an instruction to move one member of the contrast pair, followed by an instruction to move that same object or the second member of the contrast pair. The second object also varied in whether it was accented contrastively or unaccented depending on the condition. They found that when the second object contained a pitch accent the participants were more likely to anticipate (look sooner to) the previously unmentioned object. In contrast when the second object contained no pitch accent they were more likely to anticipate the previously mentioned object. The pitch accent signaled a contrast between the previously mentioned item and the currently mentioned one; the absence of a pitch accent in comparison signaled that the previously mention object was being mentioned again.

Eyetracking techniques have been used with children as well. Arnold (2008) attempted to replicate the results of Dahan et al. (2002) with 4- and 5-year-olds, but was only able to replicated the anticipatory looks to the previously mentioned object in the unaccented condition; she did not find anticipatory looks to the previously unmentioned item in the accented condition.

She found this to be true however for both her 4- to 5-year old group and her adult control group. The fact that both groups showed the same effects suggests that the absence of anticipatory fixations to the previously unmentioned item in the child group were the result of some difference in the stimuli between Dahan et al. (2002) and Arnold (2008, and not that 4- to 5-year-old children had not yet acquired an understanding of "accentation."

Researchers have noted that participants are extremely sensitive to differences in stimuli across eyetracking experiments. For example, some researchers (Weber et al., 2006; Ito & Speer, 2008) have attributed the failure of Sedivy, Tanenhaus, Chambers, & Carlson (1999) to find a facilitative effect of contrastive stress because the displays weren't sufficiently large in their experiment. Sedivy et al. (1999) used only four objects per display: two that were members of a contrast set (a pink comb and a yellow comb), a color competitor (a yellow bowl), and a distracter (a metal knife). They also allowed participants to watch as the displays were being set up, a process that took approximately 20s before the instructions even began. This extended exposure to the displays potentially allowed the participants to memorize the locations of each of the objects, thus they did not ever have to search for the target (and consequently fixate on other objects). Furthermore in this setup, referring to the bowl as "the yellow bowl" is unnecessary given there is no other bowl to contrast it with. Participants in this experiment upon hearing "Now, touch the YELLOW/yellow….," regardless of whether or not it was contrastively stressed, most likely assumed it would refer to the comb since this is the only object that must necessarily be disambiguated by color. In comparison, Ito and Speer (2008) used 11-celled displays with each cell containing three to five ornaments of the same type in multiple colors so participants in their experiment could not use color to predict the upcoming referent.

Investigating the acquisition of contrastive pitch expansion in Japanese, Ito et al. (ms) ran into similar issues with their display size. Their goal was to test 6-year-old children's understanding of contrastive pitch expansion. The Japanese language does use pitch accents, but at the word level rather than the phrasal level like English. Words in Japanese can be distinguished by their pitch accent contours, similar to the minimal stress pairs in English. For example the only difference between ka-ma ("pot") and ka'-ma ("turtle") is the placement of a pitch accent on the first syllable of "turtle." If the word ka'-ma needed to be contrastively focused in an utterance, the speaker would expand her overall pitch range. In other words the pitch accent on ka'-ma would reach an even higher F0 peak as compared to when the word was not being contrastively focused.

In their first experiment, adults and 6-year-olds were presented with a two by two grid with one animal in each of the four cells. Two of the animals matched in type but were differentiated by color (pink cat, green cat). The third cell held an animal that served as a color competitor for one member of the contrast pair (green monkey) and the final cell held a distracter (orange turtle). Participants were instructed to locate two of the animals by pressing a key (adults) or pointing (children). Instructions were always given in pairs per visual display:

6. a. Pin'ku-no ne'ko-wa doko?
   Pink cat where
   Where is the pink cat?

   b. Jaa, MI'dori-no/mi'dori-no ne'ko-wa doko?"
   Then GREEN/green cat where
   Then, where is the GREEN/green cat?

In these examples capitalization represents contrastive pitch expansion whereas lower case represents the absence of the expansion. In the critical trials the second instruction referred to the second of the contrastive pair and the color adjective either had the contrastive expansion or did not. For adults Ito et al. (ms) found the presence of the pitch expansion had a facilitative effect; adults were faster to locate the target when the adjective was emphasized as compared to when it was not. In comparison children showed no such anticipatory eye-movements. They discovered, however, that children and adults seemed to have developed different task strategies. Adults in the second instruction tended to look most at the color competitor (green monkey) initially, perhaps assuming that the green cat was too obvious of a choice to be the target. Thus the pitch expansion facilitated *recovery* from this incorrect strategy of looking at the color competitor. Children showed no such bias at the beginning of the second instruction. Given the simplicity of the visual displays, perhaps their eye-movements could not be facilitated further by the presence of the pitch expansion because they were already able to locate the target quickly even without the felicitous pitch expansion.

In their second experiment (Ito et al., ms) they increased the difficulty of the task by presenting participants with a two by three grid with each cell holding three of the same animal type in different colors. Instructions were again presented in pairs but a new condition was introduced in which the contrastive pitch expansion was produced infelicitously with a novel animal:

7. a. Pin'ku-no ne'ko-wa doko?
Pink cat where
Where is the pink cat?

b. Jaa, MI'dori-no/mi'dori-no sa'ru-wa doko?"

Then GREEN/green monkey where

Then, where is the GREEN/green monkey?

Both children and adults showed anticipatory eye-movements in the felicitous condition (pink cat -> GREEN cat) when the adjective was emphasized as compared to when it was not, although the children were slightly delayed in the execution of their fixations as compared to the adults. However, only adults were misled or *garden-pathed* into looking at the previously mentioned animal by the infelicitous use of the expansion (pink cat -> GREEN monkey). Ito et al. hypothesized that children could not be garden-pathed because their slower processing did not give them sufficient time to make use of the contrastive expansion. They noted that the children in general tended to look back to the previously mentioned animal more so than the adults. This perseveration is an example of what is referred to as the "kindergarten-path effect" in the literature (Trueswell, Sekerina, Hill, & Logrip, 1999; Snedeker & Yuan 2008). Because the children were still attending to the previous animal at the onset of the second instruction, by the time they were able to process the prominence on the adjective in the second instruction more reliable segmental information from the noun had unfolded thus overriding the information from the pitch prominence.

In a later experiment (Ito et al. 2009a & 2009b) the experimenters changed the discourse marker "jaa" used to transition between the two instructions to the longer, but semantically equivalent, "sorejaa" and also increased the length of the duration between the discourse marker and the color adjective. With these modifications Ito et al. (2009a & 2009b) found that the older half of the 6-year-old group did show garden-pathed fixations when the contrastive pitch expansion was used infelicitously like the adults did.

Contradicting previous research that investigated children's comprehension of contrastive focus (Cruttenden, 1985; Cutler & Swinney, 1987; Gualmini et al., 2002; Hornby, 1971; Wells et al., 2004), Ito et al. (ms, 2009a & 2009b) found evidence that children as young as 6 years old are able to process contrastive pitch expansion in Japanese and use it to make predictions as to the identity of the upcoming referent. Unlike earlier research which presented children with "out-of-the-blue" decontextualized sentences and asked them to make judgments about the sentences, Ito et al. (ms, 2009a & 2009b) presented children with pairs of instructions and used an eyetracker to determine if the presence of contrastive pitch expansion had an effect on children's fixation patterns as compared to the absence of contrastive pitch expansion. If the children had not been able to incorporate the information conveyed by the contrastive pitch expansion, we would have expected children to wait for the segmental information from the noun to unfold before they executed a fixation to the target. However we see instead that children look sooner to the target when the presence of felicitous contrastive pitch expansion suggests that the second animal is going to be repeated from the first. Furthermore, the older children also incorrectly fixated on the previously mentioned animal in cases where the pitch expansion was used incorrectly with a novel animal.

Given that 6-year-old Japanese-speaking children are able to use contrastive pitch expansion to make predictions about upcoming speech, we would expect that 6-year-old English-speaking children should be able to use the contrastive pitch accent to the same end. In fact, there is reason to believe that the task should be somewhat easier for English-speaking children. First, since English does not use pitch accents at the word level, there is less ambiguity as to whether a word has been contrastively focused or not. Japanese-speaking children must determine if the prominence on the first syllable of a word such as KA'-ma ("turtle") is due to

contrastive expansion or simply due to the fact that ka'-ma must be pronounced with a pitch accent on the first syllable. English-speaking children would not have to make such a judgment. Second the English equivalent to

8. "Sorejaa, MI'dori-no ne'ko-wa doko?"
   Then, GREEN cat where?

Would be something like "Now, where is the GREEN cat?" Ito et al. (ms, 2009a & 2009b) found that when the children did not have sufficient time before the onset of the critical "GREEN" they did not show garden-pathed fixations to a competitor. English-speaking children will have the benefit of four syllables ("now, where is the…") prior to the onset of the adjective ("GREEN"), which is one syllable longer than the Japanese equivalent in Ito et al. (2009a & 2009b). Thus we would expect to find a garden-path effect in English speaking children around the age of 6.

This research adopted the methodology and materials of Ito et al. for use with English-speaking children and adults. It was hypothesized that using the more sensitive eyetracking methods used by Ito et al. (ms, 2009a & 2009b) and other studies conducted with adults would reveal effects of intonation on children which previous research had been unable to discover. Namely children should be facilitated by the felicitous use of L+H* and garden-pathed by the infelicitous use of L+H*. We would also expect the children to be delayed in their fixations as compared to the adults.

Methodology

Participants

Fifty-five 6- to 7-year olds (m= 6;5) were recruited through the Developmental Language and Cognition Lab at Ohio State and through the Center Of Science and Industry

(COSI) in Columbus, Ohio. The age range was expanded from the original age range in Ito et al. (ms, 2009a & 2009b) in order to increase the chances of finding a garden-path effect since Ito et al. only found an effect in the older 6-year-old children. Three children were excluded because they did not complete the task, another was excluded because the child was non-native speaker of English, and a fifth had to be excluded due to experimenter error. All children received a small prize for participating. Adult participants were twenty-seven undergraduate students enrolled in Psychology 100 at Ohio State. Four adults had to be excluded because they were non-native speakers of English, and another had to be excluded due to experimenter error. Adults received credit towards Psychology 100 for participating in the experiment.

Visual stimuli

The visual stimuli were identical to those used by Ito et al. (ms) in experiment 2 with the exception that 12 new displays were created and added to Ito et al's original 36 displays for the purpose of obtaining more data from each individual participant. The same eight different animals types (lion, rabbit, cat, fish, frog, squirrel, turtle, and monkey) and four colors (pink, orange, purple, and green) were combined to make the 12 additional displays. A given display only contained six of the animal types: one type per each of the six cells of the display arranged in a two by three celled grid as demonstrated by the sample display in Figure 1.

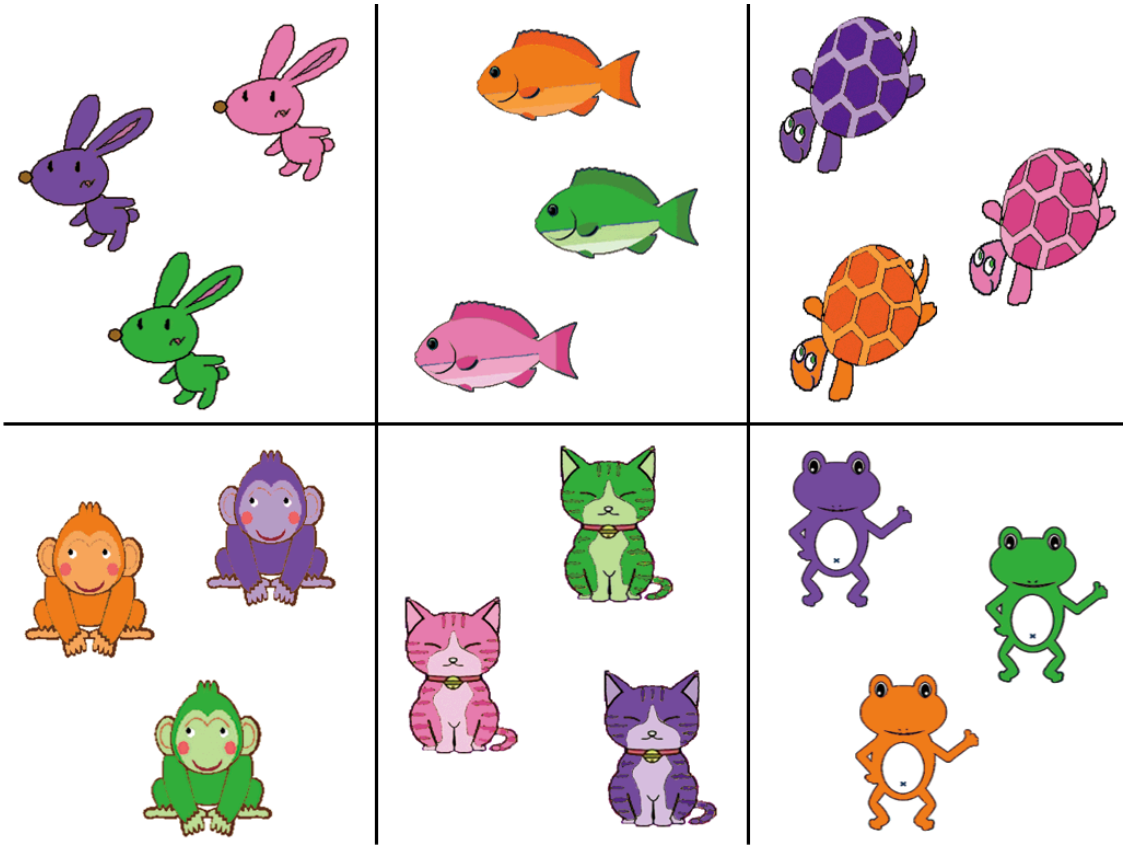

Figure 1: Sample visual display

Each cell always contained three of the same animal type that appeared in three of the four colors. The location and color combination of the animal triads was rotated for each display so that location was not predictive of animal or color combination.

Auditory Stimuli

Prompt: Where is the pink cat?			
Target Trials		Filler Trials	
1. repeat animal, L+H*	Now, where is the GREEN cat?	1. repeat color, L+H*	Now, where is the pink LION?
2. repeat animal, H*	Now, where is the green cat?	2. repeat color, H*	Now, where is the pink lion?
3. novel animal, L+H*	Now, where is the GREEN monkey?	3. novel color, L+H*	Now, where is the green LION?
4. novel animal, H*	Now, where is the green monkey?	4. novel color, H*	Now, where is the green lion?

Table 1: Sample set of utterances for all conditions

The prosodic manipulations for the auditory stimuli were adapted from the original Japanese instructions, making them similar to the English instructions used in Ito and Speer (2008). The first instruction (the prompt) was always accented with H* on the color adjective followed by !H* on the animal as shown in Figure 2.

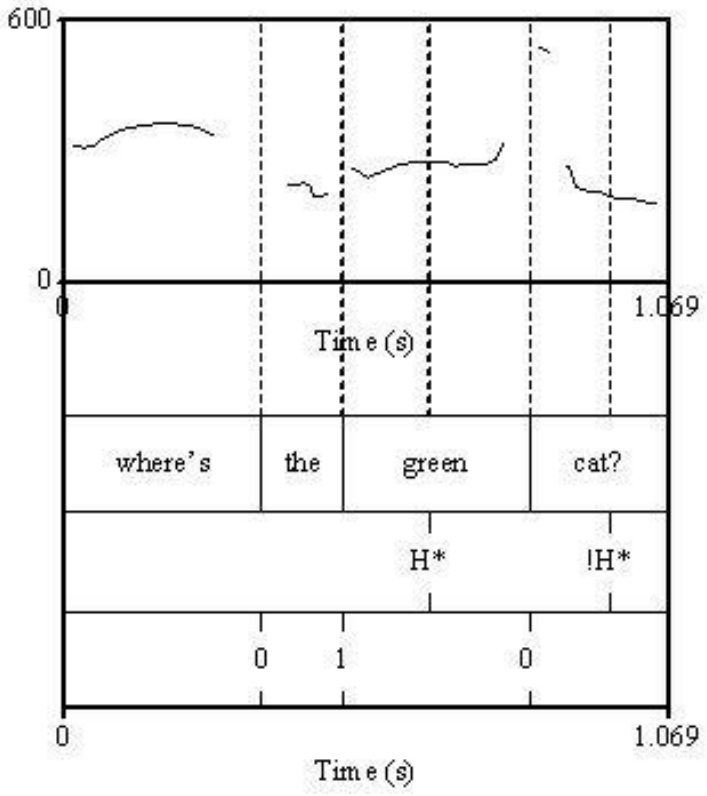

Figure 2: Sample H* !H* contour

The H* !H* pattern was chosen as the baseline for L+H* because in a piloted production version of the Christmas tree decorating task by Ito and Speer (2008), H* followed by !H* was the most common intonational pattern used by the participants in the non-contrastive ornament sequences. The second instruction (the target) varied depending upon the condition. Table 1 lists a sample utterance set for each of the conditions. In the felicitous conditions (conditions 1 and 2) the

prosody was either neutral like the prompt or the adjective was accented with contrastive L+H* followed by deaccentation on the noun as shown in Figure 3.

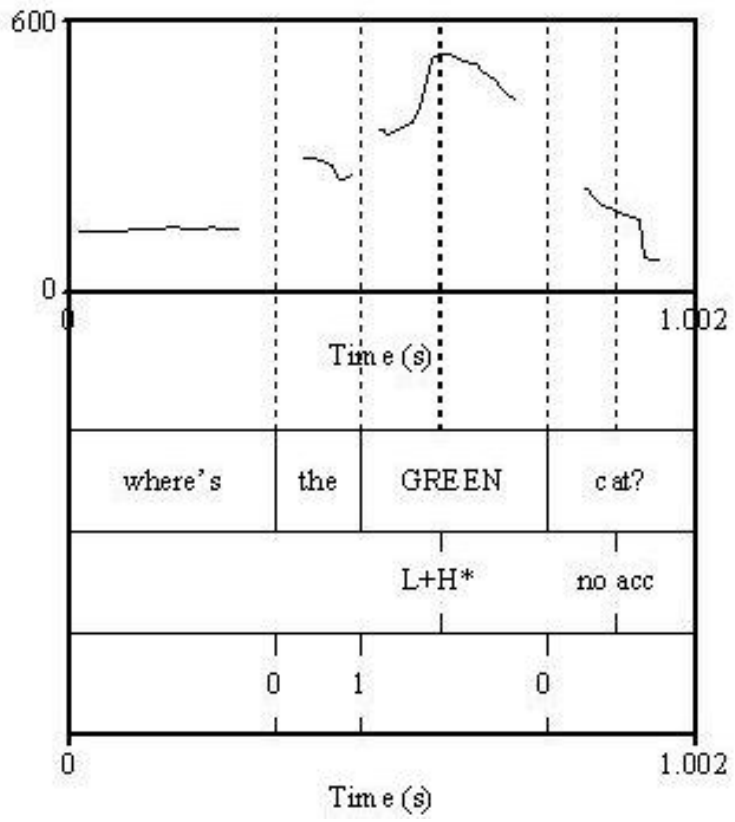

Figure 3: Sample L+H* deaccentation contour

The target animal in these conditions was always the same animal referred to by the prompt. The infelicitous conditions (conditions 3 and 4 in Table 1) followed the same intonational patterns, but the target animal was novel with respect to the prompt.

Filler trials were constructed in the same manner as the target trials except that instead of contrasting between different colored animals of the same type, the filler trials contrasted different animals that shared the same color; i.e. in a felicitous filler trial the color was repeated between the prompt and the target utterance and the critical noun (rather than the critical

adjective) was accented with L+H* (refer to Table 1). Ito and Speer (2008) showed that when displays are organized by type—as they are in these displays—L+H* on the noun has a much more delayed effect on eye-movements relative to the effects caused by L+H* on the adjective. Thus it was not expected that the filler trials would interfere with the participants' responses in the target trials.

Using these stimuli, two lists were created such that all target adjective-noun pairs that appeared with L+H* deaccentation contour in List 1 appeared with the H* !H* contour in List 2. Similarly all adjective-noun pairs that appeared with the L+H* deaccentation contour in List 2 appeared with the H* !H* contour in List 1. Thus all items were counterbalanced in terms of pitch contour across lists. The fillers were similarly counterbalanced.

All instructions were produced by a female native speaker of English and recorded using Praat with a 16-bit mono signal at a 44 kHz sampling rate. The target contours of all productions were verified by a highly trained transcriber using the ToBI labeling system (Beckman & Ayers, 1997). A pair-wise t-test was run to compare the average duration and F0 peak of the adjectives in the felicitous and infelicitous L+H* conditions to their corresponding H* conditions. Adjectives that were accented with L+H* were significantly longer and had significantly higher F0 peaks as compared to adjectives accented with H*. Table 2 shows the average durations and F0 peaks for the adjectives and nouns in the four conditions.

Averages	Adj duration (ms)	Noun duration (ms)	Adj F0 peak (Hz)	noun F0 peak (Hz)
1) L+H* repeat	**315**	360	**527**	181
2) H* repeat	295	363	278	190
3) L+H* novel	**337**	342	**523**	186
4) H* novel	308	321	268	188
Paired t-value (df = 11)	3.22**	1.11	67.72***	-1.32

Table 2: Average adjective and noun durations and F0 peaks

Procedure

Participants were seated in front of a Tobii 1750 eyetracker with a 17 inch screen. Before the experiment began they were told they would be looking at pictures of groups of animals and that they would be asked to locate some of the animals. Adults were instructed to indicate which cell contained the target animal by pressing a key on the keyboard; children were instructed to point at the screen and then the experimenter recorded the cell number with the keyboard. Once the experiment began participants were presented with a centering cross which they were required to fixate on for 1 second before a trial would begin. A trial consisted of two separate auditory instructions, a prompt and a target, which were presented with a single visual display. The prompt utterance ("Where is the pink cat?") began 1 s after a display appeared on the screen. The second instruction ("Now, where is the green cat?") began after either the adults or the experimenter (for the children) pressed one of the six numbered keys on the keyboard to indicate the answer for the prompt. The cross then reappeared after the response to the target instruction was recorded on the keyboard. The experiment proceeded in this same manner for all 48 trials except for a short break after the 24th trial.

Results

A fixation proportion for each subject was calculated for each condition by dividing the total number of looks to the target by the number of trials for each condition (in this case 6). This ratio was calculated for each time stamp, which was approximately every 20ms from the onset of a trial. These fixation proportions were then averaged into 100 ms windows (5 ratios for each window) aligning from the onset of the noun in the second instruction and then plotting forwards and backwards from the alignment point for each subject. This procedure of aligning from the onset of the noun was established by Ito and Speer (2008) and also used by Ito et al.

(ms, 2009a & 2009b). Aligning from the onset of the noun has advantages over other alignment points, such as from the onset of the adjective. Aligning from the adjective would not take into account the durational differences between various items ("pink cat" for example is much shorter than "purple turtle"), which might obscure the time course of the fixations. The point between the offset of the adjective and the onset of the noun is the only stable point across all of the items. We are also still able to investigate the effect of pitch accent on fixations during the unfolding of the adjective simply by plotting backwards from the noun, so there is no loss of data in aligning from the onset of the noun.

Four 100ms windows prior to the onset of the noun and eight 100ms windows after the onset of the noun were analyzed for all age groups for all conditions. 400ms prior to the onset of the noun was chosen as the cutoff point because the average duration of the adjectives accented with L+H* was slightly over 300 ms. The fixation proportions up to 800ms after the onset of the noun were analyzed because in one experiment which tested children's ability to use phrase breaks to interpret syntactically ambiguous utterances (Snedeker & Yuan 2008), the researchers did not see a difference in children's fixation patterns across conditions until after 700ms after the critical noun in the utterance. Thus between 700ms and 800ms after the onset of the noun was estimated to be the longest it would take for children to show an effect of pitch accent. It was predicted that adults would show an effect of pitch accent on their fixations within the first 300ms after the onset of the noun based on previous research (Ito & Speer, 2008). However the duration of the analysis windows was set to 100ms because some researchers (Snedeker & Trueswell, 2003) showed this length of window was necessary to best estimate the time course of the fixations.

To remove possible inflation of fixation proportions to the target due to subjects happening to fixate on the target before the necessary segmental information had unfolded, all trials where subjects fixated on the target at the onset of the adjective were removed from the data. This correction resulted in the removal of an additional seven adults and thirty children because they had three or fewer useable trials out of six in at least one of the conditions. Since an analysis of variance assumes that for each cell in the analysis there is an equal number of measurements and it was decided that subjects with fewer than four useable trials in a given conditions posed too great of a violation of this assumption since they had at most half the number of expected measurements. It is not surprising that more children than adults needed to be removed from the analysis since we know from previous research (Ito et al. ms, 2009a & 2009b) that children tend to continue to fixate on the previous cell during the second instruction more so than the adults. The corrected data was then analyzed using repeated measures ANOVAs by items and by subjects for each 100ms window comparing the proportion of fixations to the target in the felicitous conditions, the proportion of fixations to the competitor in the infelicitous conditions, and proportion of fixations to the target in the infelicitous conditions with pitch types as a within subjects and items factor and list as a between subjects and a within items factor. No significant effects of list were found so they will not be discussed further.

A significant effect of condition on fixation proportion was found for both age groups. Figures 4 and 5 show the fixation proportions to the target averaged across subjects for the adults and the 6- to 7-year old children respectively, comparing when critical adjective was accented with L+H* versus when the adjective was accented with H*. Eye-movement data are plotted against time, justified forwards and backwards from an alignment point coincident with the onset of the noun. The vertical lines in these figures (as well as all subsequent figures) indicate the

average onset and offset of the critical adjective and noun for both the L+H* condition (solid lines) and the H* condition (dotted lines).

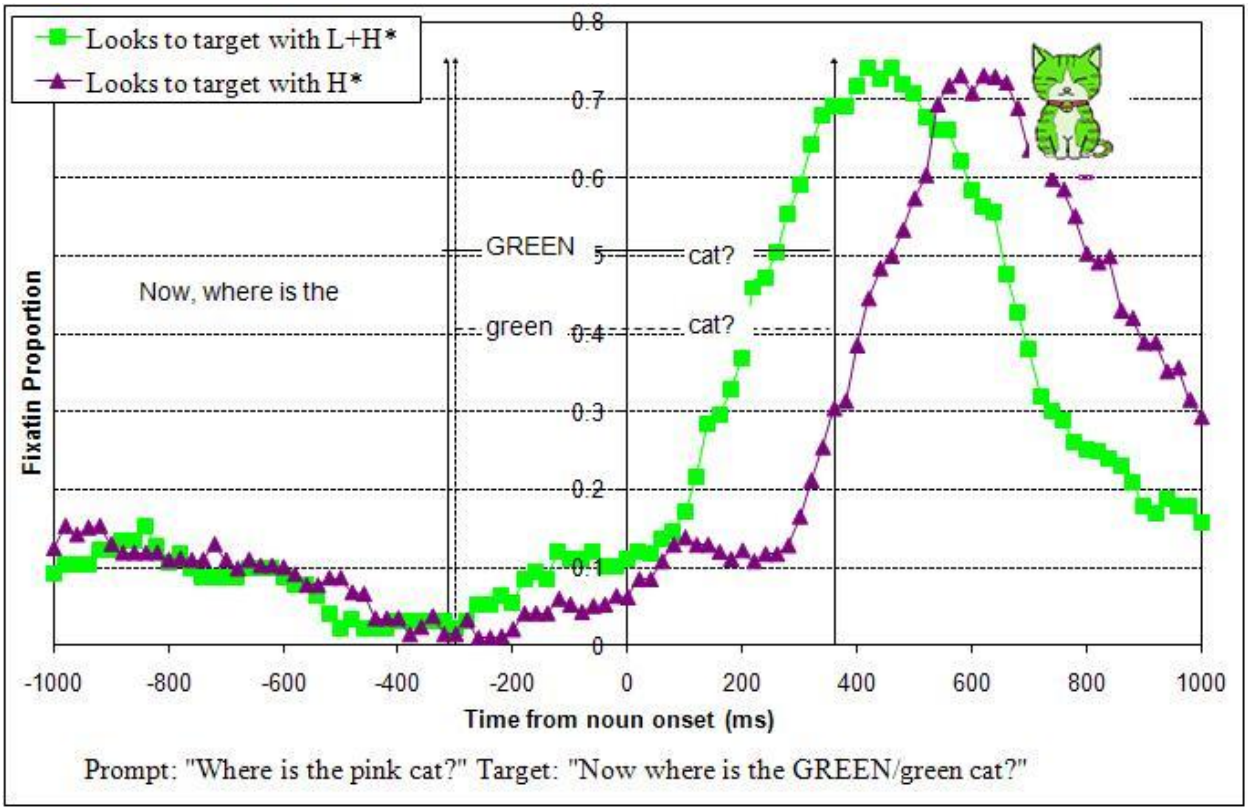

Figure 4: Adult proportion of fixations to target in felicitous conditions

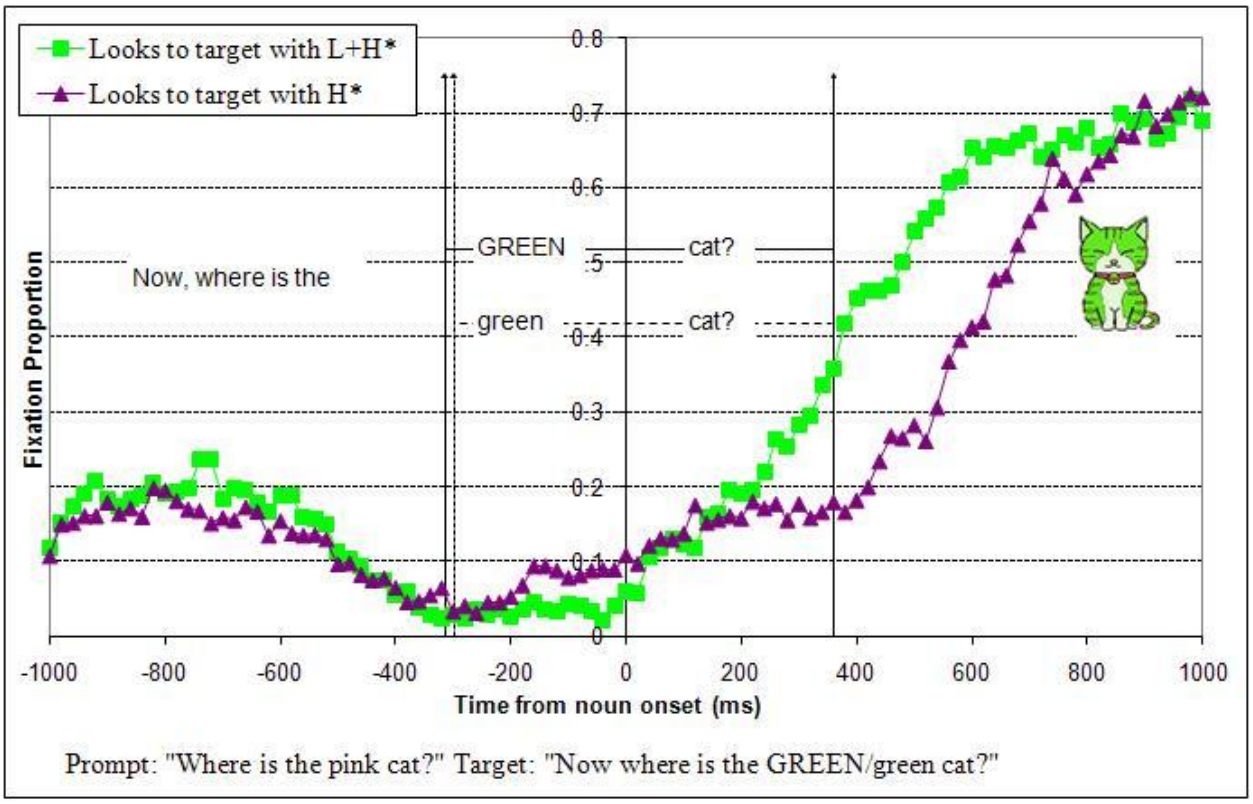

Figure 5: Child proportion of fixations to target in felicitous conditions

In the felicitous conditions both adults and children fixated on the target cell sooner when the target adjective was accented with L+H* (as in "Where is the pink cat? Now, where is the **GREEN** cat?") as compared to when it was accented with H* (as in "Where is the pink cat? Now, where is the **green** cat?") These fixation patterns suggest that the presence of L+H* had a facilitative effect on the execution of fixations to the target animal (the green cat). Table 3 shows the results from the subjects and items ANOVAs for both age groups for each time window comparing felicitous L+H* against H* on the adjective.

	F-values from ANOVAs			
	Adult		Children	
Time from noun onset	F1(1, 25)	F2(1, 11)	F1(1, 33)	F2(1, 11)
-400 to -300	0.1936	0.2385	1.8001	1.7734
-300 to -200	1.3615	4.4447	1.0145	2.9019
-200 to -100	0.9745	0.4447	5.6726*	5.1619*
-100 to 0	1.3236	0.4555	6.0543*	4.9186^
0 to 100	1.085	2.1174	0.2299	0.7151
100 to 200	19.5911***	10.2893**	0.0228	0.1512
200 to 300	54.6089***	27.5800***	1.3956	0.0997
300 to 400	53.7087***	25.7688***	11.4326**	2.8423
400 to 500	25.2816***	8.4239*	19.9293***	8.9517*
500 to 600	0.746	0.0117	26.1438***	13.6980**
600 to 700	6.1214*	3.524^	12.4488**	9.8380*
700 to 800	27.5851***	9.0945*	1.5528	1.0053
Signif. codes:	0.001 '***'	0.01 '**'	0.05 '*'	0.1 '^'

Table 3: Adult and child proportion of fixations to target, comparing felicitous L+H* and H* conditions

From this table we can see that adults first show a significant difference in fixation proportions in the 100-200 ms window. Given that it takes the average adult approximately 200 ms to execute a fixation (Allopenna, Magnuson, & Tanenhaus; 1988), the appearance of a significant difference in fixation proportions as early as 100 ms implies that adults had already made their decision about the target 100 ms before the onset of the noun's segmental information. While the children do show a significant difference in fixation proportions in the felicitous conditions, this difference is delayed as compared to the adults; the subjects analysis first reaches significance in the 300-400 ms window and items analysis reaches significance in the 400-500 ms window (refer to Table 3). The children also show significant effect in both the subjects and items analyses for the -200ms to -100ms window and also a significant effect in the subjects analysis for the -100ms to 0ms window. From Figure 5 we can see that the proportion of

fixations to the target in the H* condition are higher as compared to the fixations to the target in the felicitous L+H* condition. This pattern of fixations suggests that in the L+H* condition children were less likely to fixate on the target (the green cat) during the adjective as compared to the H* condition, which was not predicted based on previous research. If anything we would have expected the proportion of fixations to be higher in the L+H* condition during these time windows since it is possible that that participants could have been reacting to the presence of L+H* just shortly after the onset of the adjective. It is possible that this difference is an artifact of this particular group of subjects. Children are known to show greater variability in experiments as compared to adults, thus if the experiment was conducted again with a different set of 6- to 7-year-old children, this difference that we see in the 200ms prior to the onset of the noun might disappear.

In the infelicitous conditions, both adults and children showed garden-pathed looks to the previous animal when the target adjective was accented by L+H* and not when it was accented by H*. Figures 6 and 7 show the fixation proportions averaged across subjects for the adult and 6- to 7- year-old groups respectively, comparing fixations to the target and to the competitor with infelicitous L+H* versus H* plotted against time.

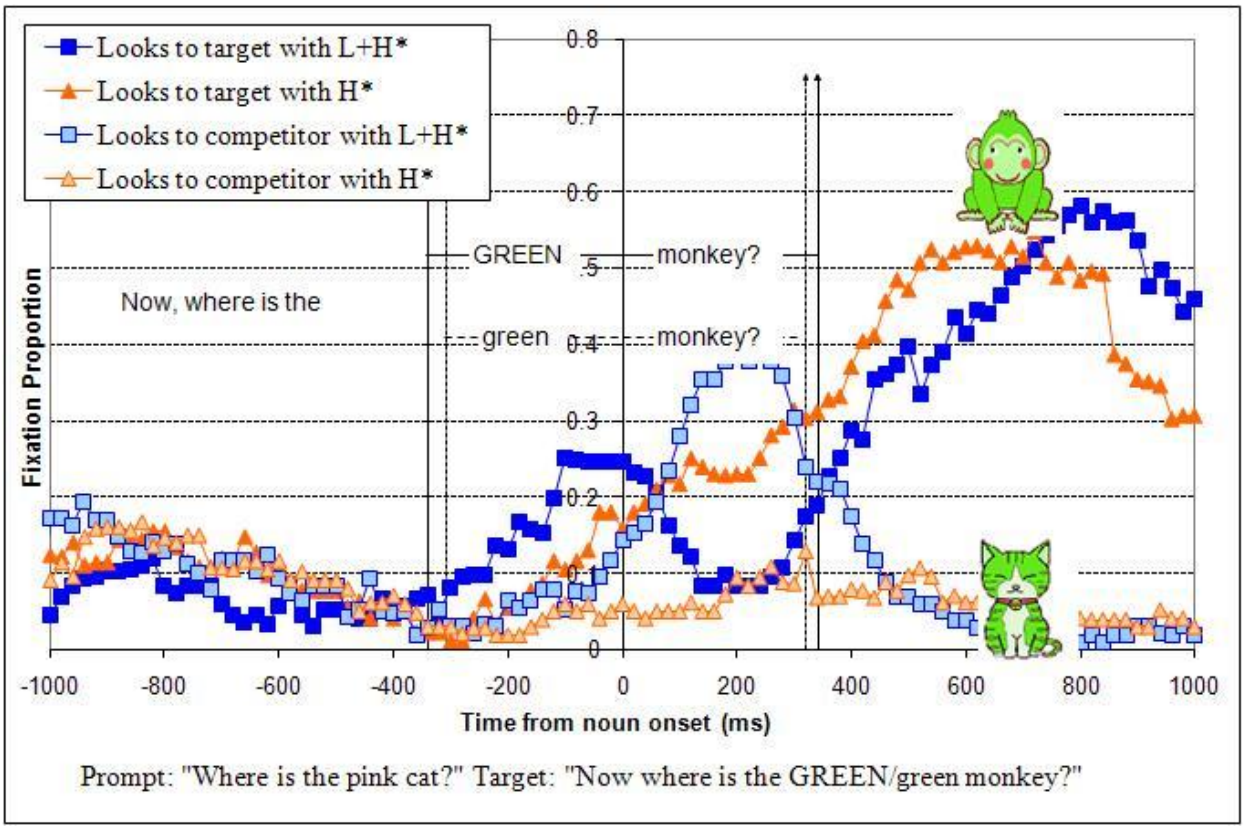

Figure 6: Adult proportion of fixations to target and competitor in infelicitous conditions

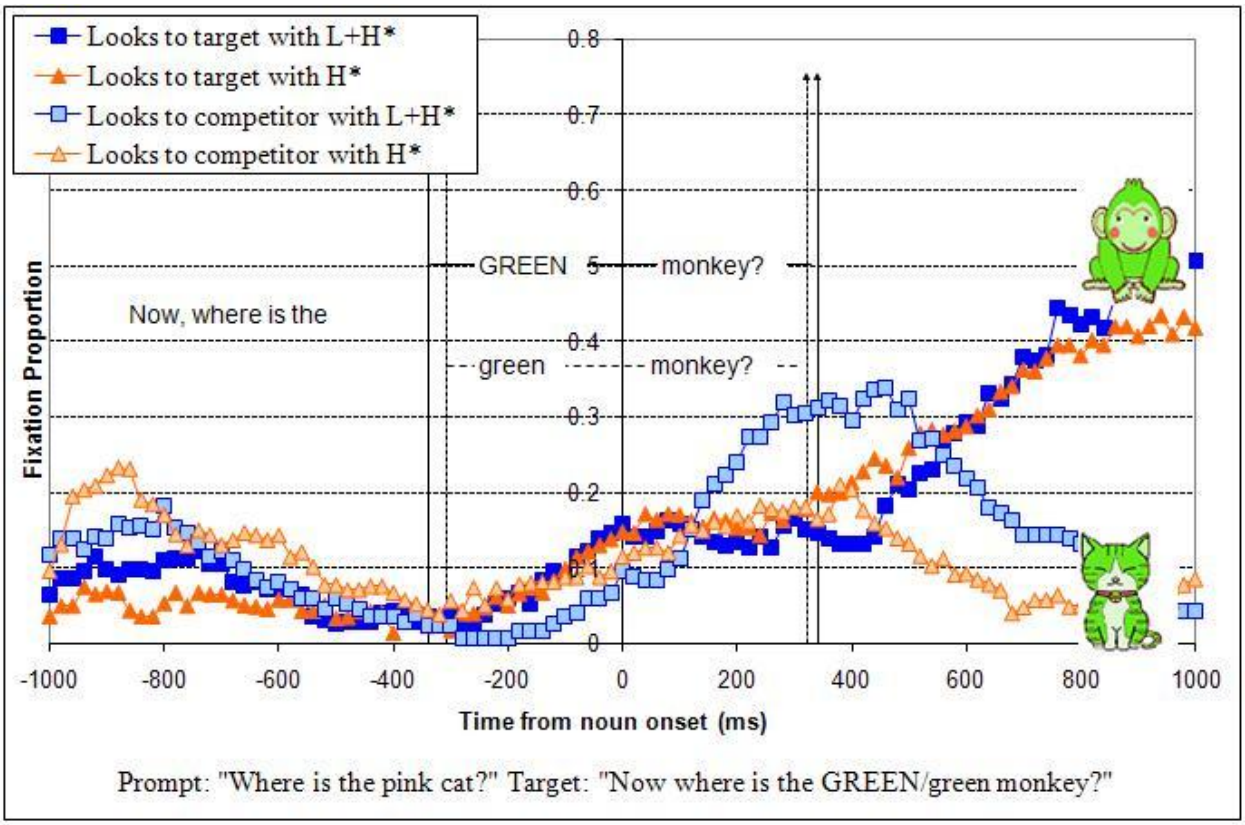

Figure 7: Child proportion of fixations to target and competitor in infelicitous conditions

When participants heard a sequence such as "Where is the pink cat? Now where is the GREEN monkey?" fixation proportions to the competitor (the cat) began to rise before the fixation proportions to the actual target (the monkey). In comparison when participants heard a sequence such as "Where is the pink cat? Now where is the green monkey?" adults hardly ever looked back to the previously mentioned animal, and children looked back significantly less as compared to when the target adjective was accented with L+H*. Table 4 shows the results of the subjects and items ANOVAs for each time window for both age groups comparing looks to the competitor given infelicitous L+H* versus H* on the adjective.

	F-values from ANOVAs			
	Adult		Children	
Time from noun onset	F1(1, 25)	F2(1, 11)	F1(1, 33)	F2(1, 11)
-400 to -300	0.1902	1.2192	0.385	0.49
-300 to -200	0.6586	0.8463	4.2529*	19.550**
-200 to -100	1.3852	5.7017*	3.6972^	5.2016*
-100 to 0	1.64	1.4487	1.0976	0.6335
0 to 100	14.6466 ***	6.1122*	0.3152	0.2261
100 to 200	32.517***	19.1389**	1.2474	0.8934
200 to 300	38.2935***	22.355***	8.6665**	4.8559^
300 to 400	20.8469***	11.3739**	6.483*	7.5898*
400 to 500	0.4224	0.1597	13.4228***	12.1876**
500 to 600	0.2984	2.1848	15.9496***	20.4056**
600 to 700	0.3085	2.8876	5.7709*	7.5123*
700 to 800	2.2319	3.4971^	4.1725*	5.7085*
Signif. codes:	0.001 '***'	0.01 '**'	0.05 '*'	0.1 ' ^'

Table 4: Adult and child proportion of fixations to competitor, comparing infelicitous L+H* and H* conditions

For adults the difference in fixation proportions (the light blue line and the light orange line in Figure 6) became significant during the 0 to 100 ms window. Children in comparison did not show a significant effect until the 200-300 ms window where the subjects ANOVA was significant and items ANOVA was marginally significant. This suggests that both adults and children responded to the presence of the L+H* before processing the segmental information from the target noun.

Similar to the felicitous conditions, in the child group we see an effect of condition in the -300m to -200ms window for both the subjects and items analyses, and a significant effect in the -200ms to -100ms window in the items analysis. From Figure 7 we can see that during this time window the fixations to the competitor in the H* condition (the light orange line) surpassed the fixations in the L+H* condition (the light blue line). Again since we have no reason to expect

why the proportion of fixations to the competitor should be significantly less in the L+H* condition versus the H* condition, it is possible that it is an artifact of this particular group of children and may disappear in a different set of children.

Based on previous research (Ito & Speer, 2008), we might have expected that fixations to the target (the green monkey) would be delayed when the adjective was accented with infelicitous L+H* as compared to when it was accented with H*. However no significant effect of pitch accent on the fixations to the target was found in any of the time windows for either age group. In the adults, the subjects analyses approached significance in the -100ms to 0ms window ($p < 0.1$), the 100ms to 200ms window ($p < 0.1$), and the 200ms to 300ms window ($p < 0.1$). We can see from Figure 6 that the fixations to the target in the L+H* condition (the dark blue line) is lower as compared to the target in the H* condition (the dark orange line) from 100ms to 300ms after the onset of the noun, which is what me would predict if we expect participants to executing a fixation based on the pitch accent after the onset of the noun and if we expect the infelicitous pitch accent to cause delays in fixations to the true target. No windows for the 6- to 7-year-old group reached significance. It is possible that although the displays used in this experiment were more complicated that those used in Arnold (2008) or Sedivy et al. (2005), they were not sufficiently complicated to cause the delays in fixations to the target that Ito & Speer (2008) found. Participants were still able to locate the target quickly even after having first fixated on the competitor.

Discussion

This research replicates the findings of Ito et al (ms, 2009a & 2009b) and also the findings of various adult studies (Weber et al., 2006; Ito & Speer, 2008). Most importantly it demonstrates that children between the ages of 6 and 7 have already acquired an understanding

of the difference between H* and L+H*, namely that L+H* evokes a contrast between the item currently under discussion and a previously mentioned item. This study also refutes previous research conducted with children that claimed children in some cases as old as 10 years old were not able to use prosodic cues in order to make judgments about speech (Cruttenden, 1985; Cutler & Swinney, 1987; Gualmini et al., 2002; Hornby, 1971; Wells et al., 2004).

It is not the case, however, that the English-speaking children were performing at the adult-levels. They showed delayed timing of the facilitation and garden-pathed effects in comparison to the adult control group. Children's delays overall in executing fixations as compared to adults has been well documented in previous literature (Trueswell et al., 1999; Arnold, 2008; Snedeker & Yuan, 2008). Since the 6- to 7-year-old children were able to use the presence of L+H* to make predictions about the identity of the upcoming referent, we know that it is not the case that they have a poor understanding of the implications of L+H* in the context of this experiment. Rather the delays must be caused by some other cognitive process still in development. Perhaps the processes necessary to execute an eye-movement are still under development during this age group. It is known, for example, that 8-year-old children require additional working memory as compared to older children in order to inhibit eye-movements (Eenshuistra, Ridderinkhof, Weidema, & van der Molen; 2006). Based on this finding we would expect that the 6- to 7-year-old children would be slower as compared to the adults to recover from their incorrect fixations to the competitor. We do in fact see this in the data: the 6- to 7-year-old children continued to fixate on the competitor significantly more in the infelicitous L+H* as compared to the H* condition from 200ms to 800ms after the onset of the noun. In comparison, the adults had finished attending to the competitor by 400ms after the onset of the

noun. It follows that as children's working memory continues to develop they should be faster in recovering from their incorrect fixations to the competitor.

While previous research (Trueswell et al., 1999; Snedeker & Yuan, 2008; Ito et al. ms) has found that children are more likely to persist in fixating on a previous item and furthermore that they have difficultly in recovering from this initial fixation (the so-called "kindergarten-path" effect) this effect does not appear to be at play in this research. In Ito et al. (ms) the kindergarten-path effect explained the absence of a garden-path effect in the children, since they discovered that the children in general tended to look back to the previously mentioned animal regardless of whether or not the adjective was contrastively emphasized. However, the children in this research do show a robust garden-path effect in the presence of infelicitous L+H*. Furthermore, once the children have been garden-pathed by the infelicitous use of L+H* they do not persist in fixating on the competitor as we might expect if they were unable to revise their initial interpretation of the instruction. We can see that in Figure 7 the proportion of fixations to the target (the green monkey) eventually do surpass the proportion of fixations to the competitor (the green cat) at approximately 600ms after the onset of the noun. It is possible that children younger than 6 might be subject to the kindergarten-path effect because their mental capacities are even less developed than 6-year-old children. It is also still an open question as to when children do become completely adult-like in their processing of these pitch accents. Further research with additional order age groups will need to be conducted to answer this question.

Conclusion

English-speaking children between the ages of 6 and 7 can comprehend contrastive L+H* and use it predictively for referent resolution in a visual search task. Children's fixations to a target are facilitated when contrastive L+H* is used felicitously—in comparison to more

"neutral" prosody—and garden-pathed when it is used infelicitously. Their fixations are also delayed as compared to adults, which suggests that the processes necessary execute fixations are still in development.

Acknowledgements

I would like to thank Kiwako Ito for giving me guidance and access to her research; Shari Speer and Laura Wagner for advising me during the course of this research; Mary Beckman and Cynthia Clopper for their much needed insight into the project; Ping Bai for assistance with data analysis; Elizabeth McCullough for assistance in creating the auditory stimuli; and Brittany Baker for assistance in creating the visual stimuli and coding the auditory stimuli.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Arnold, J. E. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*. *108*, 69-99.

Beckman, M. E. (1996).The parsing of prosody. *Language and Cognitive Processes*. *11 (1/2)*, 17-67.

Beckman, M. E. & Ayers G. M. (1997). *Guidelines for ToBI labeling*, vers 3.0 [manuscript]: Ohio State University.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. Journal of Child Language 13, 643-661.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech. 29*, 201-220.

Cutler, A. & Swinney, D. (1987). Prosody and the development of comprehension. *Journal of Child Language 12*, 643-661.

Dahan, D. Tanenhaus, M. K. & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*. *47*, 292-314.

Eenshuistra, R., Ridderinkhof, K. R., Weidema, M. A., & van der Molen, M. W. (2006). Developmental changes in oculomotor control and working-memory efficiency. *Acta Psychologica. 124*, 139-158

Gennari, S., Meroni, L. & Crain, S. (2005) Rapid relief of stress in dealing with ambiguity. In J. Trueswell and M. Tanenhaus (eds) *Processing World Situated Language: Bridging the Language-as-product and Language-as-action Traditions*. Cambridge: MIT Press.

Gualmini, A., Maciukaite, S., & Crain, S. (2002). Children's insensitivity to contrastive stress in sentences with ONLY. *In PWPL 9.1: Proceedings of the 26th Annual PLC*. University of Pennsylvania.

Hornby, P. (1971). Surface structure and the topic-comment distinction: a developmental study. *Child Development 42*, 1975-1988.

Hornby, P. & Hass, W. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research, 13*, 395-399.

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (ms.). Intonation facilitates contrast resolution: Evidence from Japanese adults & 6-year olds.

Ito, K., Jincho, N., Yamane, N., Minai, U., & Mazuka, R. (2009a). Use of emphatic prosody in Japanese adults & 6-year olds. Poster presented at The 15th Annual Conference on Architectures and Mechanisms for Language Processing, Barcelona, Spain.

Ito, K., Jincho, N., Yamane, N., Minai, U., & Mazuka, R. (2009b). Use of emphatic pitch prominence for contrast resolution: An eye-tracking study with 6-year old and adult Japanese listeners. Paper presented at Boston University Conference on Language Development 34, Boston, MA.

Ito, K. & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*. *58*, 541-573.

Ladd, D. R & Schepman, A. (2003). ''Sagging transitions'' between high pitch accents in English: experimental evidence. *Journal of Phonetics. 31*, 81-112.

MacWhinney, B. & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behavior. 17,* 539-558.

Pierrehumbert, J. B. (1980). The phonology and phonetics of English intonation. PhD dissertation. Massachusetts Institute of Technology.

Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack. *Intentions in communication*. pp. 271-312. MIT Press.

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109–147.

Snedeker, J. & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48,* 103-130.

Snedecker, J. & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language. 58,* 574-608.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science. 268,* 1632-1634.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*. 73, 89-134.

Trueswell, J. C. & Tanenhaus, M. K. (2005). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. In J. C. Trueswell, and M. K. Tanenhaus. *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*. pp. 3-37. MIT Press

Watson, D. G., Tanenhaus M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science. 32,* 1232-1244.

Weber, A. Braun, B. & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents." *Language and Speech*. *49*, 2006, 367-392.

Wells, B., Peppé, S. & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31* 749-778.